DOCUMENT RESUME

ED 208 819 IR 009 415

AUTHOR Tamura, Kiyoko
TITLE JAPAN/MARC: Its Present Status and Future Prospects.
PUB DATE Aug 80
NOTE 18p.; Paper presented at the Annual Conference of the International Federation of Library Associations (Manila, Philippines, August 18-23, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cataloging; *Databases; Flow Charts; Foreign Countries; Information Seeking; Library Automation; *Library Catalogs; National Libraries; *Online Systems
IDENTIFIERS Japanese National Bibliographic Information System; *JAPAN MARC; *Kanji Script

ABSTRACT

The creation of Japan/MARC--a system employing Kanji script (Chinese characters as adopted by the Japanese)--as the online database for Japanese bibliographic records marks a significant development in the availability of the Japanese national bibliography and should give stimulus to other countries using Chinese as a predominant language. JAPAN/MARC records allow entries transcribed in Kana and/or the roman alphabet of author, title, subject heading, and other access points. In addition to the core database, the National Diet Library is establishing data files by subjects, authors, series, and publishers, and expects to include retrospective records back to 1969, eventually including serials and journal articles. The text provides detailed information on formats, records, and procedural rules and appendices include the character set, tag relationships between UNIMARC and JAPAN/MARC, bibliographical data elements, and a flowchart of the Japanese National Bibliographic Information System. (RAA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED208819

IFLA General Conference 1980

Division: Management and Technology

Section: Information Technology

57/INF/3E

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

JAPAN/MARC; ITS PRESENT STATUS AND FUTURE PROSPECTS

By

Kiyoko TAMURA
Computer Application Section
National Diet Library
TOKYO
Japan

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

A.L. van Wesemael

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR009415

1. Introduction

MARC as defined as a national bibliography in machine readable form is a basic tool for achieving IFLA's goal of UBC (Universal Bibliographic Control). Accordingly, the development of a national MARC is the important responsibility of each national bibliographic agency internally as well as internationally.

The National Diet Library (NDL) has been preparing for publication the Japanese National Bibliography over the past 32 years of its existence, and it is the most important one NDL should prepare in pursuance of the legal provision. Equally important is the compilation of the Japanese Periodicals Index, quarterly, an index of articles appearing in the Japanese representative journals of research value in all fields of human endeavor.

To date, both have been offered mainly in book-form and catalog card, and in due course of time it is duly expected that records contained therein be made available also in machine-readable form and on-line.

2. Development of JAPAN/MARC

The full development of JAPAN/MARC is a pressing task to be performed by NDL as the producer of the Japanese National Bibliography. In its efforts to improve this bibliography, due consideration is given to its format, comprehensiveness in coverage and prompter availability. As a vehicle to implement these, the NDL set up in 1978 a Deliberative Council on JAPAN/MARC with its membership comprising representatives from public, university and special libraries, government agencies, information service, publishing and communications industries as well as people of experience and expertise in and outside the Library. The Council met several times, first, to finalize the format of JAPAN/MARC, and its conclusion was submitted for approval to the National Diet Librarian in the autumn of 1979. Consequently, it is duly expected

that distribution of MARC tapes be able to be made toward the end of 1981.

What makes JAPAN/MARC characteristic is that its main data elements should be recorded in Japanese script with kanji (Chinese characters as adopted by Japanes) as the major component. In utilizing JAPAN/MARC, technical development of both hardware and software is essential to process kanji records. To this end, 6 institutions including some university libraries in the coutry are now experimenting with test tapes from the data base. On the other hand, quite a few Japanese manufactureres are actively engaged in technical development of Japanese-character processing system, focusing on standardization of Chinese character sets, development of software for kanji processing and production of kanji input/output machines at reasonable cost. Now that the formulation of a system for use of Japanese-language bibliographic records has no longer been a tricky job, availability of JAPAN/MARC would certainly bring a great benefit to potential users.

3. JAPAN MARC Format

3-1 Prerequisites

1) The coverage of data to be offered is that of the data base being built up by NDL.
2) Any user sector can use this data base.
3) The copyright to the data will be held by NDL, who normally will, however, not make any compensational claim upon users of the data bases.

3-2 Cataloging rules

The Nippon Cataloging Rules in use by NDL conforms to ISBD, but does not follow it strictly in every detail, reflecting Japanese special circumstances.

3-3 Structure of bibliographic records

The structure of bibliographic records shall be based on ISO-2709 (Documentation - Format for bibliographic informa-

tion interchange on magnetic tape). In consideration of the physical character of magnetic tape, each dat will be recorded on 9 track tape with record density of 1600 BPI.

3-4 Data arrangement of the bibliographic record (See Appendix 2 & 3)

As regards the arrangement of various elements in a given bibliographic record, UNIMARC (Universal MARC Format) prepared by IFLA should be taken into due consideration. This format is recommended as the one for recording information in Roman alphabet. However, it is questionable for this format to be adopted as it exactly is for Japanese-language records. It is because the Japanese-language record must be indicated in kanji and for its readings in romanized form or kana (phonetic Japanese alphabets).

3-5 Codes used

The character set used for JAPAN/MARC is represented by codes of both alphanumeric character (1 byte) and kanji (2 bytes).

1) Codes of alphanumeric characters are based on ASCII or EBCDIC.

ISO-646 (7-bit coded character set for information processing interchange)
JIS-C6220 (Code for information interchange)
EBCDIC

2) For kanji codes, the standard codes of JIS (Japan Industrial Standard) are used. (See Appendix 1)

JIS-C6226 (Code of the Japanese graphic character set for information interchange)
JIS-C6225 (Code of the control character set for Japanese graphic characters for information interchange)

3-6 ISBN

In Japan a system of the Japanese Book Code has been in use since 1970 for distribution and sales of commercial publications with code numbers assigned and printed by

individual publishers. Publications with such code numbers printed are approximately 70% of the total production, while one-third of the total publishers are using this code system.

Recently, however, the Japan Publishers Association has decided to adopt ISBN in consideration of the growing popularity of ISBN abroad and of the future prospects of the Japanese publishing industry. As a result, some publishers began to use ISBN in April 1980 and the great majority of the publishers are expectd to follow suit in early 1981.

The following is an example of bibliographic records of JAPAN/MARC.

4. Processing of kanji records

4-1 Characteristics of kanji

1) There are quite a number of characters in use. Characters to be used vary in frequency from one field or subject to another.

Examples:

Substantially large Chinese-Japanese dictionaries	49,964
Characters used for NDL printed cards	c.10,000
JIS character codes	6,349
NDL character codes	4,567
Characters used for magazines	c.3,300
Characters used for newspapers	c.3,600

2) The structure of a character is often complicated, and the same character has in many cases variant forms.

Example:

KEN (Sword): 剣 劍 劒 劔 劍

3) There are many different words with identical sound but with different characters.

Example:

KISHA: 汽車 (train) 記者 (reporter)

4) The average ratio of a Japanese word to an English word in terms of the number of letters is 1:2.5.

Example: 川 : River (1:5)

5) The Japanese sentence consists of Kanji (Chinese characters) and Kana (phonetic symbols). The average number of Kanji used in a sentence is 30 to 40%.

6) The same character has two or more different readings. One way is on-yomi (readings taken from Chinese) and the other kun-yomi (native Japanese readings).

Example: 剣 : on-yomi → KEN kun-yomi → TSURUGI

7) There are more than two ways of filing. In general, the filing of entries is based on the proper reading of their

Chinese characters.

The reading is transcribed in the following ways:

a. Kana

b. romanization by the government system (kunrei-shiki)

c. romanization by the Hepburn system

8) In the Japanese sentence, there is no word-for-word division. Only punctuation is used.

Of these charactersitics, 7) and 8) are especially noteworthy. As shown in 7), the reading of kanji must be transcribed in kana and/or romanized form as an indispensable means for filing and access point.

4-2 Development of multi-use systems

We using the Japanese language as the native language should assume the pioneering task of developing a system for the processing of kanji-written information. What has been done in Japan over the past decade was to intorduce the computer devised and developed in the United States and then to devise a new system that has enabled th alphanumeric code to handle kana.

Recently, computer technology in Japan has been advanced to such an extent that quite a few manufactureres are now producing and developing machines and systems that are capable of handling kanji as well as the alphabet and kana.

4-3 Bibliographic information systems at NDL

The National Diet Library installed computer in 1971 with immediate objective of processing Japanese-language bibliographic records, and since then developed some systems, each designed to meet the specific requirements in compiling various bibliographies previously prepared and published on manual operation.

These haveing been accomplished, NDL's efforts have now been turned to the compilation, storage and distribution of national bibliographic records in the form of JAPAN/MARC. The basic requirements to this end were to develop kanji-processing technique and equip such hardwares as kanji

input machines, high-speed kanji printers and kanji display terminals.

The data base of JAPAN/MARC started to be built up in 1977 and the following year saw the first production of a weekly list and a quarterly index of the Japanese National Bibliography through automated compilation. For the flow chart of the system, see Appendix 4.

The coverage of the data base is as follows:

1) The average input per year is approximately 40,000 records, of which official publications, both central and local, take up about 17%. Since the beginning of 1980, the data base includes pamphlets, monographic series, European-language publications issued in Japan and micro-form publications. It is estimated that eventually the data base will cover about 80% of a total of Japanese domestic publications excluding serials.

2) The coverage by subject is as follows:

Subject	%
Politics. Law. Administration	6.2
Parliamentary publications. Legal materials	0.8
Economics. Industries	13.6
Society. Labor	5.2
Education	4.4
History. Geography	9.9
Philosophy. Religion	3.3
The arts. Language. Literature	19.8
Science and Technology	12.8
Learning in general. Journalism. Library Bibliographies	2.6
Rare books. Old books	0.5
Special materials	2.0
Children's books	6.8
Limited-cataloging materials	11.5
Others	0.6

3) The coverage by year of publication is as follows:

Publication year	Data stored in the years		
	1977	1978	1979
1979-	0%	(1 item)	65%
1978	(4 items)	71%	30%
1977	81%	22%	2%
1976	14%	2%	0.7%
1975	2%	1%	0.3%
pre-1975	3%	4%	2%

Note: Publications printed in a particular year and one year before occupy the greatest percentage of the data stored in that year.

4-4 On-line information retrieval service

Well over 10 institutions in Japan are providing on-line information service using the data bases produced overseas and/or by themselves. Of them, only a few institutions are currently dealing with Japanese-language data in their efforts toward including such data in their proper service. To porvide Japanese-language data service on-line, the prerequisite is that a user should be equipped with a kanji terminal that can meet the following requirements:

1) The terminal should not be too expensive. The cost will range from $4,800 to $8,000.

2) The terminal can be used for multiple systems for information retrieval.

3) The terminal should be capable of handling both alphanumeric and kanji data.

Fortunately, the formulation of kanji codes by JIS (Japan Industrial Standard) has been a big step forward, while an advance in LSI (large scale integration) technology has led to gradual cost-down of the equipment. On the other hand, a protable-type kanji terminal using public telecommunications network, less expensive in cost, is under development and will be put to practical use some time in 1981.

The information system under development is that a kanji-display terminal will receive on its image screen specific kanji information transmitted from the data base by means of inputting key works in roman alphabet or kana. The next step will be to use kanji for key works at the terminal. Key works will cover title, author, subject, classification number, ISBN, NBN (National Bibliographic Number, etc.

5. Conclusion

The creation of JAPAN/MARC as the data base for Japanese-language bibliographic records in machine-readable form will mark a significant development not only in prompt availability of Japanese National Bibliography but also in giving stimulus and some examples to other countries where Chinese is a predominant language.

In terms of data description, JAPAN/MARC should have in each record elements that show the readings transcribed in kana and/or roman alphabet of author, title, subject heading information that serve as access point, and this factor makes JAPAN/MARC unique and different from any other MARC in existence.

In addition to the core data base, the NDL is in the process of establishing such particular data files as for subjects, authors, monographic series and publishers.

Also, our data base is to be expanded to include retrospective data of some 220,000 books published for 1969-1977.

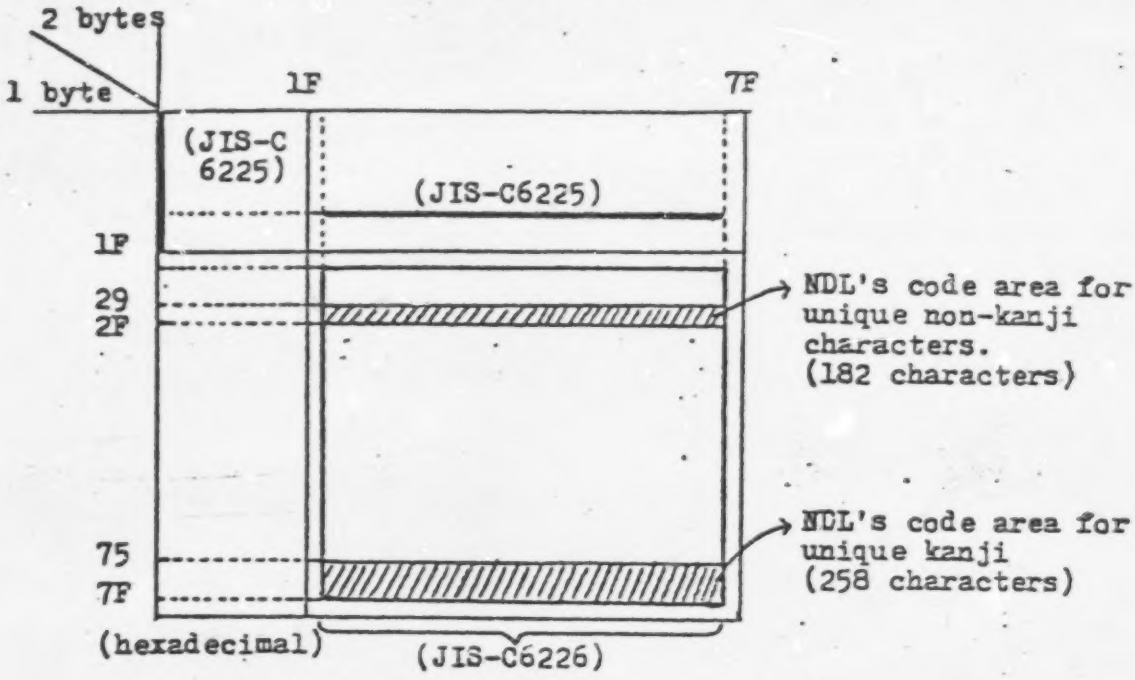
As a further step to go, the NDL will begin to study the possibility that bibliographic records of serials and journal articles be stored in and made available from a data base.

Thus, our goal is to put JAPAN/MARC to multiple use not only for prompt availability of cataloging information and eventual preparation of a union catalog, but also as an efficient aid for such works as selection, reference and inter-library loan.

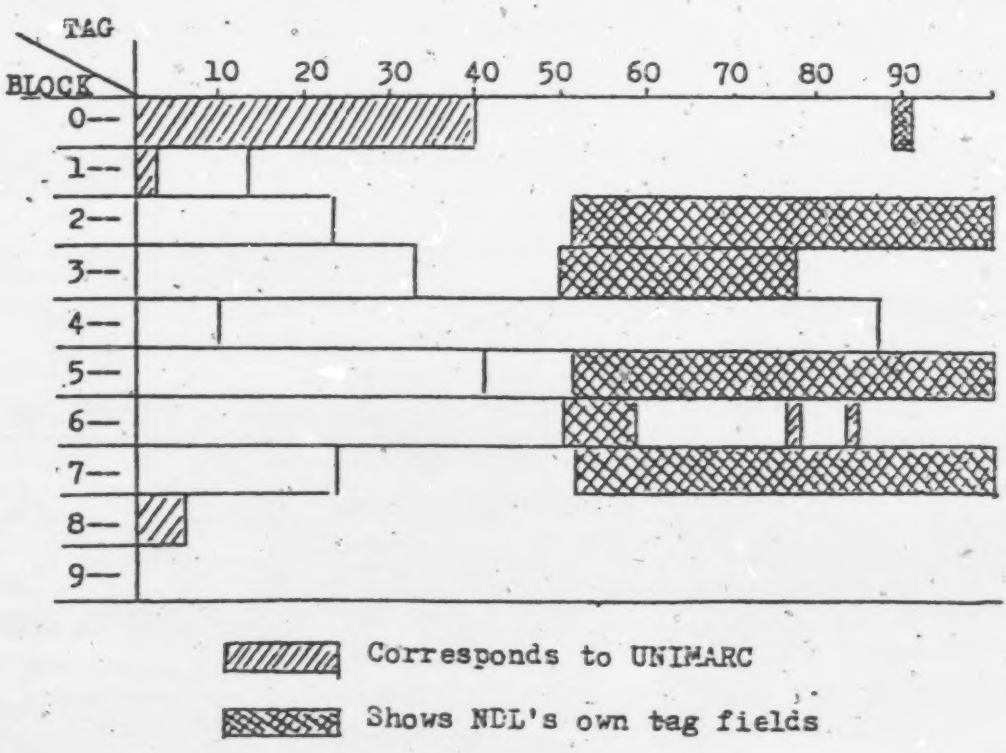
Building up a kanji data base requires a joint work of many catalogers and a big amount of money. The comprehensiveness of coverage is another essential factor. No

other library can well afford to do this. In this sense, JAPAN/MARC will bring inmeasurable benefits to other libraries in the nation in facilitating their automation work and the formation of a cooperative network of libraries sharing the same data base. Sooner or later, foreign libraries, especially with substantial collection of Japanese publications, will eventually enjoy the usefullness of our MARC data.

(Appendix 1) Character Set



(Appendix 2) Tag Relationship between UNIMARC and JAPAN/MARC (Monographs)



(Appendix 3) Bibliographic Data Elements

BLOCK	TAG	FIELD	SUBFIELD ID.	SUBFIELD
ification	001	Record Identifier Field		
	010	ISBN		Same as UNIMARC
	020	National Bibliography No.		
Information	100	General Processing Data		Same as UNIMARC
	101	Language of the Work		
ctive mation	251-	Title and Statement of Authorship	$a	Title proper
			$b	Other Title Information
			$d	Volume Numbering
			$f	Statement of responsibility
	265	Edition	$a	Edition statement
	270	Publication, etc.	$a	Place of publication, etc.
			$b	Name of publication, etc.
			$d	Date of publication, etc.
	275	Physical Description	$a	Extent of item
			$b	Dimensions
	280	Series	$a	Series statement
			$b	Volume Numbering
			$d	Name of part
			$f	Numbering within part

13C15C

	291-	Title and statement of Authorship (single volume-level)	$a $b $d $f	Same as 251
Notes	350	Notes	$a	General Note
	354	"	$a	Original Title of translation
	360	"	$a $b	Binding Price
	377	"	$a	Contents Note
Related Title	551-	Romanized (or kana) form of Title proper	{$a $x}	Kana Romanized form
	580	Romanized (or kana) form of Series Title		
	591-	Romanized (or kana) form of Title (single volume-level)		
Subject Analysis	650	Personal Name used as Subject Heading	{$a $x $b}	Kana Romanized form Kanji
	658	General Subject Headings		
	677	*NDC	{$a}	Number
	685	NDL Classification		
Intellectual Responsibility	751-	Author (Title proper)	{$a $x $b}	Kana Romanized form Kanji
	791-	Author (single volume-level)		

1970 (5)

(Appendix 4) Japanese National Bibliographic Information System (Monographs)